

Workflow Reflection

What was the most challenging part of the workflow? Why?

The most challenging part of the workflow was the data preprocessing phase. This included handling missing values, standardizing inconsistent formats, and dealing with imbalanced data. It required a lot of attention to detail because errors in this stage can lead to poor model performance or biased outcomes. Ensuring the data was clean, representative, and fair added an extra layer of complexity.

How would you improve your approach with more time/resources?

With additional time or resources, I would build a more automated and modular preprocessing pipeline using tools like Python scripts or dataflow tools (e.g., Apache Airflow). I would also invest time into more thorough exploratory data analysis (EDA) and consult with domain experts to identify meaningful features. Access to larger, more diverse datasets would also help improve the model's generalizability and fairness.